Remote Sensing of Environment 130 (2013) 39–50

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data

J. Christopher Brown [a,*], Jude H. Kastens [b], Alexandre Camargo Coutinho [c], Daniel de Castro Victoria [d], Christopher R. Bishop [b]

[a] Department of Geography, University of Kansas, 1475 Jayhawk Blvd., 223, Lawrence, KS 66045, United States
[b] Kansas Applied Remote Sensing Program, University of Kansas, 2101 Constant Ave., Lawrence, KS 66047, United States
[c] Embrapa Informática Agropecuária, Av. André Tosello, 209, CP 6041, Campinas, SP 13083-886, Brazil
[d] Embrapa Monitoramento por Satélite, Av. Soldado Passarinho, 303, Fazenda Chapadao, Campinas, SP 13070-115, Brazil

ARTICLE INFO

Article history:
Received 17 February 2012
Received in revised form 7 November 2012
Accepted 8 November 2012
Available online xxxx

Keywords:
Brazil
Cotton
Cross Validation
Decision Tree
Land Cover
Soybean

ABSTRACT

MODIS 250-m NDVI and EVI datasets are now regularly used to classify regional-scale agricultural land-use practices in many different regions of the globe, especially in the state of Mato Grosso, Brazil, where rapid land-use change due to agricultural development has attracted considerable interest from researchers and policy makers. Variation exists in which MODIS datasets are used, how they are processed for analysis, and what ground reference data are used. Moreover, various land-use/land-cover classes are ultimately resolved, and as yet, crop-specific classifications (e.g. soy–corn vs. soy–cotton double crop) have not been reported in the literature, favoring instead generalized classes such as single vs. double crop. The objective of this study is to present a rigorous multiyear evaluation of the applicability of time-series MODIS 250-m VI data for crop classification in Mato Grosso, Brazil. This study shows progress toward more refined crop-specific classification, but some grouping of crop classes remains necessary. It employs a farm field polygon-based ground reference dataset that is unprecedented in spatial and temporal coverage for the state, consisting of 2003 annual field site samples representing 415 unique field sites and five crop years (2005–2009). This allows for creation of a dataset containing "best-case" or "pure" pixels, which we used to test class separability in a multiyear cross validation framework applied to boosted decision tree classifiers trained on MODIS data subjected to different pre-processing treatments. Reflecting the agricultural landscape of Mato Grosso as a whole, cropping practices represented in the ground reference dataset largely involved soybeans, and soy-based classes (primarily double crop 'soy-commercial' and single crop 'soy-cover') dominated the analysis along with cotton and pasture. With respect to the MODIS data treatments, the best results were obtained using date-of-acquisition interpolation of the 16-day composite VI time series and outlier point screening, for which five-year out-of-sample accuracies were consistently near or above 80% and Kappa values were above 0.60. It is evident that while much additional research is required to fully and reliably differentiate more specific crop classes, particular groupings of cropping strategies are separable and useful for a number of applications, including studies of agricultural intensification and extensification in this region of the world.

1. Introduction

For over a decade, MODIS datasets have been used for regional scale studies of agricultural landscapes. This is especially evident in studies in the Brazilian Amazon. Researchers have taken advantage of MODIS's high-temporal, moderate-spatial resolution characteristics to map crop classes at the farm field-level, allowing for the tracking of land-use and land-cover change and evaluating the implications for carbon biogeochemical cycling (Galford et al., 2010a, 2010b, 2011), deforestation trends (Anderson et al., 2005; Barona et al., 2010; Clark et al., 2010; DeFries et al., 2008; Morton et al., 2006), cropping frequency changes (number of crops per year) (Brown, et al., 2007a; Coutinho et al., 2011; Epiphanio et al., 2010; Galford et al., 2008; Jasinski et al., 2005; Martinelli et al., 2010), and effectiveness of agri-environmental governance systems (Rudorff et al., 2011). There seems to be little doubt that MODIS data are useful for these exercises, and with each passing year the possibility increases of completing studies of inter-annual changes over longer periods of time.

There is substantial variation in how researchers have used MODIS data for the purposes of classification and subsequent analysis of inter-annual change. The raw datasets may range in pixel size (250 m to

* Corresponding author. Tel.: +1 785 864 5543.
E-mail addresses: jcbrown2@ku.edu (J.C. Brown), jkastens@ku.edu (J.H. Kastens), alex@cnptia.embrapa.br (A.C. Coutinho), daniel@cnpm.embrapa.br (D.C. Victoria), bish305@ku.edu (C.R. Bishop).

0034-4257/$ – see front matter © 2012 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.rse.2012.11.009

1 km), in the vegetation index (VI) used (NDVI vs. EVI), in data preparation and pre-processing (e.g. smoothing, filtering, interpolation), and in modeling method (e.g. maximum likelihood, decision trees). Moreover, the field data used to train classification models range from rapid field surveys along roads (Epiphanio et al., 2010; Morton et al., 2005, 2006), examination of detailed farm records in particular regions (Galford et al., 2008), interviews of farmers (Arvor et al., 2008; Brown et al., 2007a; Jonathan et al., 2008), and visual interpretation of high resolution imagery (Clark et al., 2010). The ways data are processed for analysis are even more diverse, with most aiming to reduce noise in the data caused by atmospheric interference, cloud cover, sensor issues, or inaccuracies of field data collection. Many studies transform or otherwise smooth datasets with the intent of improving resulting classification accuracies (e.g. Galford et al., 2008).

With the continued stream of data beyond the projected years of the MODIS Terra and Aqua satellite platforms – they were originally designed to last 6 years when first launched in 1999 and 2002, respectively – and the recent launch of the NPOESS Preparatory Project (NPP) and eventual launch of the Joint Polar Satellite System 1 (JPSS-1), it is clear that satellite-based land change studies will rely on time-series analysis for many years to come. With the increasing use of these data to support market-driven efforts for environmental protection and climate change mitigation (e.g. Reducing Emissions from Deforestation and Forest Degradation (REDD), the Amazon Soy Moratorium, payment programs for ecological services), researchers have an important role to play in ensuring that land-use/land-cover (LULC) classifications are as accurate as possible and produced at the lowest cost (DeFries et al., 2008). Researchers must also be assured that field data are collected in a standardized fashion, ideally covering several years in order to assess annual variability and to allow rigorous out-year, out-of-sample validation that leads to the most robust analysis of, and conclusions made about, land change in Brazil's rapidly developing agricultural landscape.

While numerous authors cited above have tested the suitability of MODIS for their purposes (including the present authors), the objective of this study is to present a uniquely rigorous multiyear evaluation of the applicability of the time-series MODIS 250 m NDVI and EVI datasets for crop classification in Mato Grosso, Brazil. It is based on the pioneering work of Wardlow et al. (2007), which tested separability of crop classes in the state of Kansas based on a farm-field polygon dataset, allowing for the selection of what can be termed “best-case” or “pure” pixels from the polygons for analysis. The present study's ground reference data are unprecedented in their spatial and temporal coverage for Mato Grosso, including 415 farm field polygons across the state's main growing regions and data from 5 crop years (2005–2009). The present study assesses the separability of VI values for the crop classes from the field data, and it provides an empirically justified grouping of classes for the purposes of eventual map classification. The study also reports the level of classification accuracies potentially achievable under ideal circumstances when mapping agriculture in Mato Grosso using time-series MODIS VI data and boosted decision tree models.

2. Study area

All of the field data for this study come from the state of Mato Grosso, Brazil, a major center of mechanized agricultural production within Brazil's Legal Amazon, a bio-administrative unit (Fig. 1). Mato Grosso covers approximately 900,000 km^2 and is bordered by Bolivia to the southwest. The southern part of the state is a tropical wetland known as the Pantanal (61,726 km^2). In the north are the humid forests of Amazonia (481,129 km^2). The central part of the state is dominated by vast tropical savannas known as cerrado (360,008 km^2). The region's climate (Köppen Aw) is hot, semi-humid to humid, with pronounced seasonality marked by a dry winter season from May through October. The annual rainfall ranges from 1300 to 2300 mm. Many of the state's

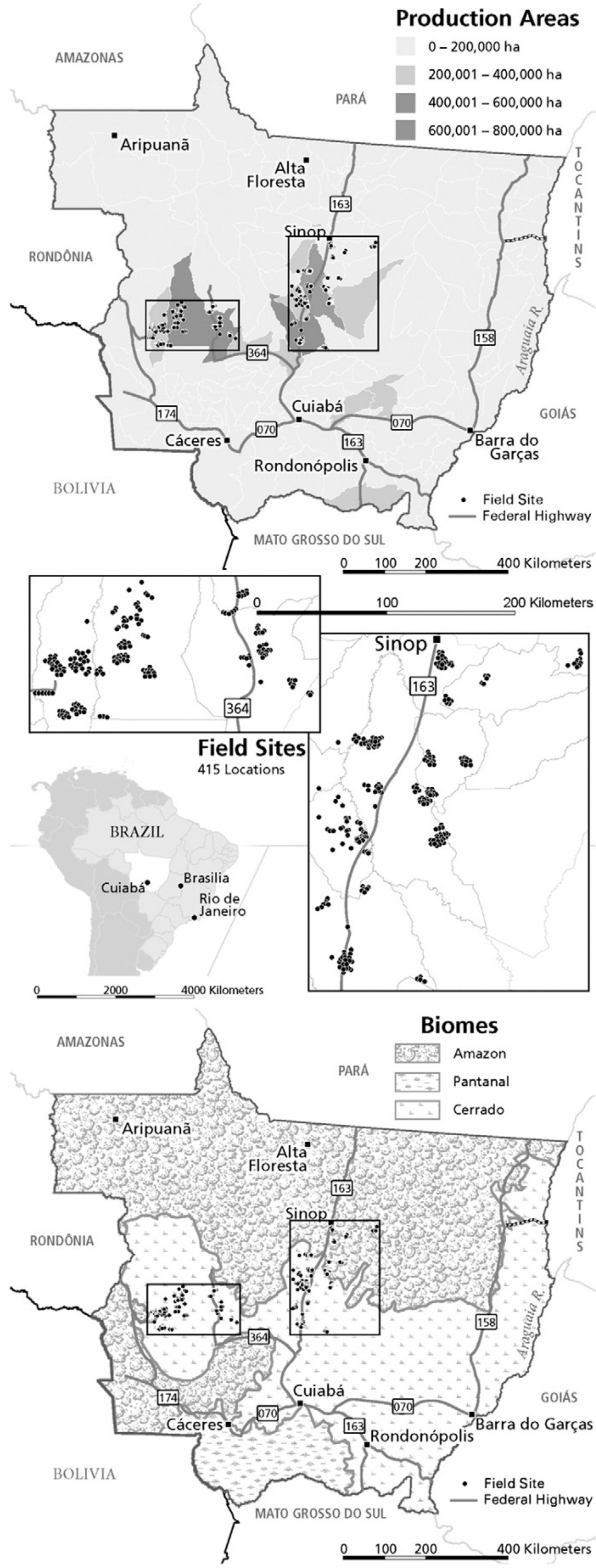

Fig. 1. Study area in Mato Grosso, Brazil. (Biome source map: IBGE).

soils, especially in the cerrado region, are old, deep, and naturally nutrient poor. Due to inputs of adequate fertilizer and lime, new breeds of crops (especially soybeans), and favorable world markets, Mato Grosso has come to be called a major "breadbasket" in a country that is quickly becoming an agricultural super-power. Of the Brazilian states, Mato Grosso leads in soybean production and is second in corn production. Mato Grosso is also a major cotton producing state, accounting for approximately half of Brazil's total production. From 2000 to 2005, land area planted with soybeans, the state's principal crop, increased at an average rate of 19.4% per year (Jasinski et al., 2005). Between 1990 and 2010, total planted area in Mato Grosso increased by a factor of 386% (from 2.43 million to 9.38 million ha), with the majority of this expansion taking place in the cerrado biome (IBGE, 2011). By 2000, double cropping soy with soy led in part to the development and spread of soybean rust. To alleviate this problem, second crop (*safrinha*) corn was introduced, and soy followed by second crop corn has become the dominant double cropping practice. In recent years, other *safrinha* crops such as sorghum and sunflowers have become popular as well (Soybean & Corn Advisor, Inc., 2011). The major crops in the region include soybeans, corn, and cotton, with soybeans accounting for nearly 70% of the total planted area during 2005–2009 (IBGE, 2011).

The field data for the study were collected within an area extending from (59° 25′ 14″ W, 14° 2′ 39″ S) [lower left] to (54° 25′ 19″ W, 11° 42′ 16″ S) [upper right], from the following 14 municipalities in the most intensely cropped region of central Mato Grosso: Brasnorte, Campo Novo do Parecis, Campos de Júlio, Diamantino, Ipiranga do Norte, Lucas do Rio Verde, Nova Mutum, Santa Carmem, Sapezal, Sinop, Sorriso, Tapurah, União do Sul, and Vera. On average across the 2005–2009 crop years, the total cropland area in these municipalities accounted for 43.5% of the total agricultural area in the state (IBGE, 2011). Most of these areas are located along the BR-163 highway that extends from Mato Grosso's capital, Cuiabá, north to Santarém, Pará on the Amazon River. A mild southeast to northwest precipitation gradient exists across the study area, with average annual precipitation ranging from 1750 mm to 1925 mm.

3. Data and methods

Two datasets were essential to this study: field data containing 5 years of agricultural land use records for hundreds of fields in Mato Grosso, and corresponding MODIS time-series satellite vegetation index data. Boosted decision tree models were used for the several different classifications that were examined. Distributional similarities between crop classes with respect to their spectral profiles resulted in the examination of three different classification schemes. In an effort to maximize model accuracy, three levels of signal processing were considered along with three levels of data filtering.

3.1. MODIS data

Sixteen-day composite Terra MODIS 250-m NDVI and EVI data from the MOD13Q1 Vegetation Indices product line (Collection 5) were used for this study. These data were obtained from the United States Geological Survey's Land Processes Distributed Active Archive Center (LP DAAC). One MODIS tile (h12v10) was required, which covered all of the field sites. Because the Mato Grosso crop calendar runs from composite period 14 (Jul 28–Aug 12; season start) to composite period 13 (Jul 12–Jul 27; season end), the analysis dataset spanned five crop years (2004 period 14–2009 period 13) and consisted of 115 scenes (5 years $*$ 23 scenes/year). Corresponding acquisition date information also was extracted from the MOD13Q1 data files, giving the specific day of the 16-day composite interval from which each MODIS pixel VI value was obtained (the same date applies to both NDVI and EVI). MODIS Quality Assurance information was not used, which simplified processing by avoiding altogether any subjective decision on how to use those data (i.e., how to drop values out, and possibly how to replace them). This allowed for analysis of the MODIS VI data in their most readily available form.

3.2. Field data

In situ data were collected via farmer/farm manager interviews in September 2009 across a wide swath of central and western Mato Grosso in central Brazil (Fig. 1). The cropping practices were recorded for individual field sites (polygons) for the 2005–2009 crop years (Coutinho et al., 2011). The field-level information was integrated into a GIS as specific polygons with attribute data that were used to explore the MODIS time-series data. Interviews proceeded after obtaining oral consent to participate in the research project, following the protocol outlined by Institutional Review Board Guidelines of the University of Kansas. A total of 40 farmers or farm managers were interviewed as research participants.

To obtain field data, authors Coutinho and Victoria presented participants with a Landsat TM image at a scale of 1:100,000 containing the areas they were responsible for farming. Each field, or *talhão*, indicated by the participant was a polygon in which cropping practices and management were purported to be homogenous throughout a given agricultural year (roughly August through July). These fields were outlined on the paper image with a marker. Fields for the 2008–2009 agricultural year were marked first, followed by fields for 2007–2008 and earlier as time, memory, or record keeping allowed. If participants expressed doubt about their ability to recognize a field boundary or report accurate planting and cropping practice information, that information was discarded. Each field received a unique identifier, and for each field the following data were collected for each agricultural year: type(s) of crop(s) and planting sequence, and (if applicable) type of fallow (i.e., true fallow or planting a particular cover crop). Each field outline was then digitized on-screen over a Landsat TM image using Quantum GIS (http://www.qgis.org/) to create a field polygon coverage. Approximately 20% of the polygons are within the humid forest biome, and 80% are within the cerrado. For simplicity, we refer to each crop year using the year in which it ends, so that the 2008–2009 crop year is now just referred to as 2009, and so on.

The resulting field polygon layer contained 415 polygons, each with one to five years of LULC information. Of 2075 potential annual field site samples, 72 records were absent, resulting in a total of 2003 annual field site samples. The fields range in size from 23 to 2793 ha with a median field size of 176 ha. The field polygons were converted to raster format corresponding with the MODIS pixel grid, and border pixels were removed so that only pixels completely interior to the field boundaries were retained. Following Wardlow et al. (2007), a single, high quality, centrally located 250-m MODIS pixel was manually selected from each polygon to represent that field site in the dataset for analysis. This approach was taken because our main objective was to assess potential class separability. As such, this method minimizes the influence of mixed pixels on the analysis (i.e., maximizes the "purity" of the spectral signatures) and ensures that each field site has equal representation in the analyzed dataset. The quality was assessed by examining the NDVI and EVI profiles of pixels centrally located in the field. Each centrally located pixel was compared to its surrounding in-field neighbors (up to 8 pixels), and the pixel with the most visual continuity with its neighbors was selected. In some of the larger, more convex fields, multiple pixels were available that met our selection criteria. Even in these instances, a single pixel was selected to maintain consistency of sampling and consistency with Wardlow et al. (2007). In instances where fields were irregularly shaped, this process was applied to the broadest, most convex part of the field. In many cases, the centroid pixel was used when found to be centrally located and of high quality. After pixel selection, the corresponding pixel-level time-series NDVI, EVI, and date-of-acquisition data were extracted from the MODIS time-series stacks to create the raw VI analysis dataset.

Nineteen unique classes are represented among the 2003 annual field site samples (Table 1). Imposing a minimum sample size of 30 immediately eliminated seven classes (amounting to 104 total samples) from the analysis (four of these classes – clearing, forest, cerrado, and reforest – would have been excluded anyway because the focus of this research is on agricultural land cover). One additional class (rice), which had the next smallest sample size at 36, was also eliminated from consideration. Rice is frequently opportunistically planted as a temporary transition crop on deforested land to help condition the ground for future soybean plantings. Consequently, rice phenology in the study area exhibits exceptionally inconsistent timing and growth, as well as irregular pre-crop VI values depending on the time of clearing (Brown et al., 2007a). This variability was reflected in the field site NDVI database, as rice samples demonstrated the largest aggregated period-by-period coefficient of variation when averaged across the 23 sixteen-day time periods of the MODIS data. Following these exclusions, 11 classes remained, amounting to 1863 annual field site samples (Table 1) and approximately 120,000 ha of agricultural land represented annually (Table 2). On average across the five study years, the crops represented in these 11 classes accounted for 91.5% of reported agricultural land area in Mato Grosso (IBGE, 2011). Fig. 2 shows median profiles and data bands from the six largest classes.

3.3. Decision tree classifiers and model evaluation methods

The commercial decision tree (DT) classifier See5 was used to perform the classifications. DTs currently serve as the main classification models for prominent national and global-scale LULC mapping efforts such as the USGS NLCD (Homer et al., 2004), the MODIS Land Cover Type product (MOD12Q1) (Friedl et al., 2002), and the USDA NASS Cropland Data Layer Program (Boryan et al., 2011). DTs are non-parametric, hierarchical classifiers that predict class membership by recursively partitioning data sets into increasingly homogeneous, mutually exclusive subsets via a branched system of data splits (Breiman et al., 1984). Key components of DTs are internal nodes (branching points), terminal nodes (end nodes or leaves), and branches (connections linking two nodes). At each internal node, the optimal independent variable and threshold value are identified that result in the best possible data split based on statistical deviance (Wardlow & Egbert, 2007). Once the DT's classification structure is established, each observation (pixel) from the dataset to which the DT is applied is passed through the tree and assigned to the class of the leaf node into which it falls.

Table 1
Number of field site samples, by year and by crop type. The 11 largest classes were used for the analysis, accounting for a total of 1863 field site samples.

Crop	2005	2006	2007	2008	2009	Total
Soy–corn[a]	101	98	130	151	156	636
Soy–millet[b]	105	102	86	88	77	458
Soy[b]	81	101	40	28	36	286
Pasture	22	21	24	22	25	114
Soy–cotton[c]	24	21	25	12	11	93
Cotton	7	15	9	18	13	62
Soy–sunflower[a]	6	6	8	14	14	48
Soy–sorghum[a]	6	6	9	14	10	45
Soy–corn–pasture[a]	4	2	11	17	8	42
Soy–pasture[b]	3	2	9	13	14	41
Soy–beans[a]	5	5	6	7	15	38
Rice	5	8	8	7	8	36
Corn	3	12	1	7	3	26
Cotton–millet	6	6	6	2	2	22
Soy–rattlepod	1	1	3	4	8	17
Forest	5	3	5	3	1	17
Cerrado	2	2	2	2	1	9
Reforest	1	1	1	1	4	8
Clearing	2	0	2	0	1	5
Total	389	412	385	410	407	2003

[a] Soy-Com class in all scenarios.
[b] Soy-Cov class in all scenarios.
[c] Soy-Com class in 4- and 2-class scenarios.

Table 2
Field site area in hectares, by year.

2005	2006	2007	2008	2009
120,209	114,760	120,541	118,317	120,619

Unconstrained DTs can be constructed large enough to fit any training dataset to any degree of accuracy when there are no training samples that have identical predictor values but different dependent variable values. Consequently, constraints must be imposed to limit tree size and mitigate overfitting. In See5, this is accomplished through the use of two parameters, the *minimum leaf size* and the *certainty factor* (CF).

To implement the minimum leaf size, See5 halts splitting when the optimal split of the training data subset at a parent node results in at least one child node that contains fewer data points than the minimum leaf size. When this occurs, that parent node is not split, and instead becomes a terminal node. The default minimum leaf size value in See5 is 2 cases. We chose to use this value for all of our analyses after extensive out-of-sample testing indicated this value to be optimal for our classification problem. Specifically, we tested minimum leaf size values of 2, 5, and 10 in a factorial design using CF values of 0%, 1%, 2%, 5%, 10%, and 25% (results not shown).

The CF dictates a form of error-based pruning, a key feature of DT development designed to mitigate overfitting and make the tree more parsimonious, so that the tree's predictive ability is more robust when applied to unseen data. Use of pruning is common in DT classification processes for LULC mapping applications (DeFries & Chan, 2000; Friedl & Brodley, 1997; Friedl et al., 2002; Hansen et al., 1996; Homer et al., 2004). Pruning involves removing parts of the tree (splits) that are expected to have a relatively high error rate or contribute little to reducing the deviance in the training data.

In See5, the CF determines statistical confidence limits for the predicted number of errors across the leaves below a test node and compares this to the predicted number of errors at the test node if it were a leaf (which is estimated from the observed number of errors at the test node). If the predicted number of errors at the test node is less than the sum of the upper limit predicted numbers of errors across the child nodes, then the leaves are pruned. Consequently, the lower the CF value, the wider the confidence interval and the more likely that pruning will occur. After testing multiple CF values, we chose to use a CF value of 1% because this value exhibited the least overfitting but also did not underfit the data (which was always the case when using a CF value of 0%). It is likely that the optimal CF value is between 0% and 1%, but the See5 software only allows integer percentage values for the CF.

Boosting is another feature of DT modeling that generates several classifiers (decision trees) rather than a single classifier, in an effort to improve classification accuracy. In See5, boosting optimizes multiple classifiers using a base classification algorithm in an iterative fashion while systematically varying the training sample to emphasize difficult-to-classify cases from previous iterations. The final 'boosted' classification output is produced by a weighted voting scheme across the multiple classifiers (RuleQuest Research, 2012). Any number of iterations can be performed, but traditionally 10 iterations have been used for most previous LULC mapping efforts where boosting was employed (DeFries & Chan, 2000; Friedl et al., 1999; McIver & Friedl, 2001; Wardlow & Egbert, 2007). Following these studies, we used boosting with 10 iterations for all of our DT analyses.

Three different approaches to the DT modeling process were explored to gain a better understanding of DT classification error. Specifically, two

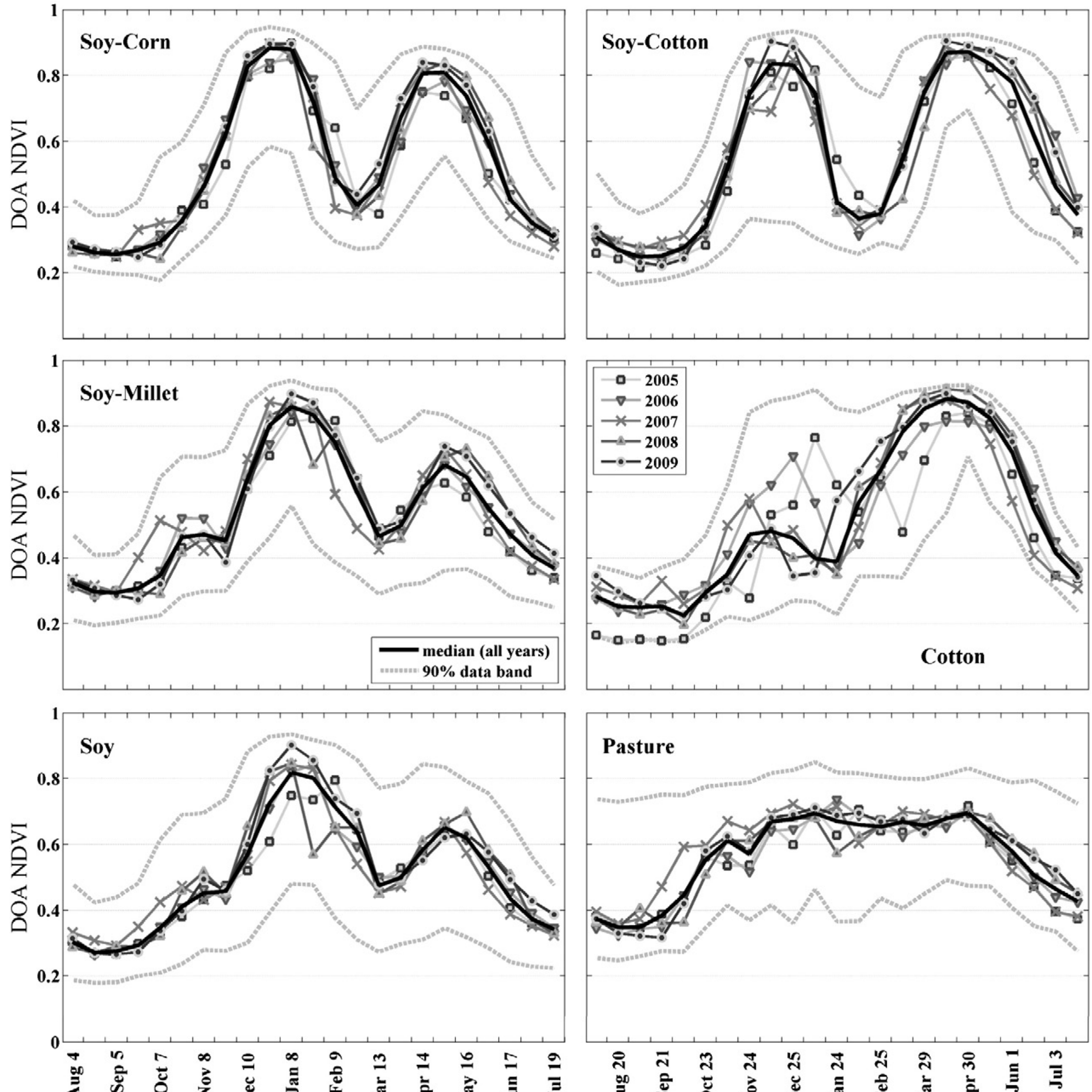

Fig. 2. Field site annual and overall median DOA NDVI profiles, along with 90% data bands, for the six largest crop type classes in the ground reference dataset. Note the similarities and differences among the four classes involving soy. Also note the wide range of NDVI values in the cotton profiles prior to the cotton signal visible in the latter part of the crop year.

out-of-sample cross validation (CV) approaches and one in-sample approach were evaluated.

For the first (and most rigorous) out-of-sample method (denoted CVYR), data from one of the five years were withheld from modeling, and the model generated using the remaining four years of data was subsequently applied to the withheld year to determine classification accuracy. This was repeated with each of the five study years acting as the withheld year, and the results were aggregated to obtain an overall accuracy estimate. As with any CV-based model error estimation exercise, one should be mindful of potential "data framing" effects that occur when the withheld data are not independent of the model training data, imparting a favorable bias to the results. Determining hold-out sets based on crop year (which are largely independent of one another) should mitigate this bias.

For the second out-of-sample method (denoted CV20), the data were pooled across all five years and were split using a stratified (by class) random sample with 80% allocated for training and 20% for validation. These data splits were created independent of sample year. The 80% subsets from each class were merged and used for model construction, and the 20% subsets were merged and used for model evaluation. Thirty independent classification runs were performed using thirty different stratified random draws of training and validation data, and validation accuracies for the thirty runs were averaged to obtain a robust overall accuracy estimate reflective of this model evaluation approach.

Though CV20 is an out-of-sample evaluation method with roughly the same holdout set size as CVYR, it is not as rigorous as CVYR. This is because data samples from different years have greater independence from one another than samples taken from the same year due to similarities in phenology (expressed through VI values) caused by large-scale spatial similarities in single-year weather patterns that heavily influence crop management and growth. Comparing results from CVYR and CV20 helps expose bias in out-of-sample validation methods such as CV20 when applied to multiyear datasets, if the intent is to develop a model that can be applied to data observations from years that are not part of the study.

For the in-sample method (denoted INSAMP), all samples were used for both model construction and model evaluation. This approach is assured to produce biased results that exaggerate a model's ability to make predictions using another dataset. By comparing INSAMP accuracies to those observed using the CV approaches, the magnitude of this bias can be examined to better understand the degree to which the model is overfitting the data.

3.4. Combining classes for DT modeling

Phenological similarities in MODIS VI profiles between different crop types can preclude clean separation during classifier development. In this situation, combining similar classes is necessary to produce robust classification models (Arvor et al., 2008; Brown et al., 2007a). Comparing VI profiles from the 11 classes in the analysis dataset using correlation and Jeffries–Matusita (JM) distance (Richards & Jia, 1999; results not shown), two logical groupings became apparent, which we refer to as 'soy-cover' and 'soy-commercial'. These two super-classes accounted for eight of the 11 cropping practices. The 'soy-cover' (Soy-Cov) class subsumed three original classes: soy, soy–millet, and soy–pasture. For these component classes, soy is the only commercial crop (occasionally *safrinha* millet is harvested, but millet is predominantly used as a cover crop in this setting). Five classes comprised the 'soy-commercial' (Soy-Com) class: soy–corn, soy–sunflower, soy–sorghum, soy–corn–pasture, and soy–beans. All of these component classes have soy followed by a second commercial crop. The remaining three classes (cotton, pasture, soy–cotton) were assessed to be sufficiently distributionally distinct to warrant individual representation in the reduced class list.

To test the general suitability of this 5-class grouping, we subjected the VI data to K-means clustering, setting the "number of groups" parameter (K) to five. The purpose was to examine to what degree the five data clusters resulting from this unsupervised classification reflected the 5-class specification defined above. The K-means algorithm used random seeding for initial clustering, so 30 replicates were evaluated. Results from the trial with the lowest sum of squared distances between individual samples and their corresponding cluster centers fairly well corroborate our 5-class scheme (Table 3). Each cluster containing at least 20% of the points from a given class is emphasized in the table. Specifically, we note that Class 1 is roughly a soy-cover class, Class 2 is a cotton class, Class 3 is a pasture class, Class 4 is a soy-commercial class, and Class 5 is a combination soy-commercial/soy-cover class. The most notable exception is that no exclusive class emerged for soy–cotton, which largely (63%) landed in Class 4. As a consequence of this observation, in addition to the 5-class specification, we also evaluate a 4-class specification, whereby soy–cotton is absorbed into the soy-commercial class (which is appropriate because *safrinha* cotton is a commercial crop in this case). As a final test to examine just how well we can distinguish soy-commercial from soy-cover, we also perform 2-class analyses whereby we exclude cotton and pasture samples and include soy–cotton in the soy-commercial class.

3.5. Data treatment variations

All analyses performed using NDVI were repeated using EVI. In total, nine different data treatments were examined for each VI dataset: three levels of signal processing were considered, with three levels of data filtering then evaluated for each signal processing level.

Table 3
NDVI DOA K-Means clustering results. Emphasized entries indicate more than 20% of a particular crop sample set was assigned to a particular class.

Crop	Class 1	Class 2	Class 3	Class 4	Class 5
Soy–corn	43	10	22	***322***	***239***
Soy–millet	***204***	7	52	36	***159***
Soy	***153***	11	22	32	***68***
Pasture	***25***	5	***75***	4	5
Soy–cotton	1	12	7	***59***	14
Cotton	4	***41***	1	8	8
Soy–sunflower	8	6	0	***11***	***23***
Soy–pasture	***13***	0	***11***	4	***13***
Soy–sorghum	***9***	0	3	5	***28***
Soy–corn–pasture	2	0	6	***23***	***11***
Soy–beans	4	2	6	***18***	***8***

The first level of processing ("RAW") used raw VI values extracted directly from the 16-day MODIS composite images and stacked in sequence to form 16-day time series. The second level of processing ("DOA") used 16-day VI time series extracted from linearly interpolated daily data developed from the pixel-level, date-of-acquisition (DOA) information that accompanied the MODIS composite data. To create the interpolated VI values, we associated each raw VI value with its DOA and simulated all missing daily VI values using linear interpolation between each pair of successive observations. Each field site was processed individually, because each has its own unique set of DOA values that is applicable to both NDVI and EVI. The length of an interpolated span can be 0–30 days, depending on which days from the two consecutive 16-day intervals that the DOA values occur (e.g., day 16 in period 1 and day 1 in period 2 leads to 0 interpolated days between those two VI values, whereas day 1 in period 1 and day 16 in period 2 leads to 30 interpolated days). Upon completion of the daily interpolations, we extracted the VI values at 'day 8' from each 16-day composite interval to produce the DOA NDVI and DOA EVI time series. The third level of processing ("SM-DOA") was similar to the second level, except that a flat-bottom smoothing algorithm (Wardlow et al., 2006) was applied to the raw time series prior to date-of-acquisition interpolation. In this algorithm, each local minimum VI value is identified in the time series, and its value is increased to the level of its lowest neighboring (i.e., immediately preceding or immediately following) VI value. This smoothing algorithm was selected because of its simplicity and its emphasis on tampering only with local minima of VI profiles, which are the most likely location for substantive data value errors when the data are developed using maximum value compositing techniques (which characterizes the 16-day MODIS VI products). See Fig. 3(a) for an example showing NDVI profiles from all three processing levels for one cotton field site sample.

The first level of data filtering ("NOFILT"), which involved no filtering, used all of the data. The second ("85/15") and third ("80/20") levels of filtering used identical filtering procedures but with a different filter stringency parameter value. The purpose was to remove outliers to purify the sample and make the results better reflect best-case classification scenarios, thereby exposing *potential* classification accuracies. First, all of the VI data across the five study years were grouped by class (which varied across the three different classification schemes examined), and period-by-period VI values were sorted. For each of the 23 annual MODIS periods, class-specific VI value percentiles were empirically determined to identify lower and upper period-specific VI tail threshold values. For the 85/15 dataset, the 15th and 85th percentiles were determined (likewise, 20th and 80th percentiles were used for the 80/20 dataset). Each annual spectral profile in a class was examined to count the number of tail occurrences among the 23 comprising periods. Finally, any profile with more than half of its VI values (i.e., 12 or more data points) occurring in the appropriate class- and period-specific distribution tails was excluded from the dataset. To illustrate

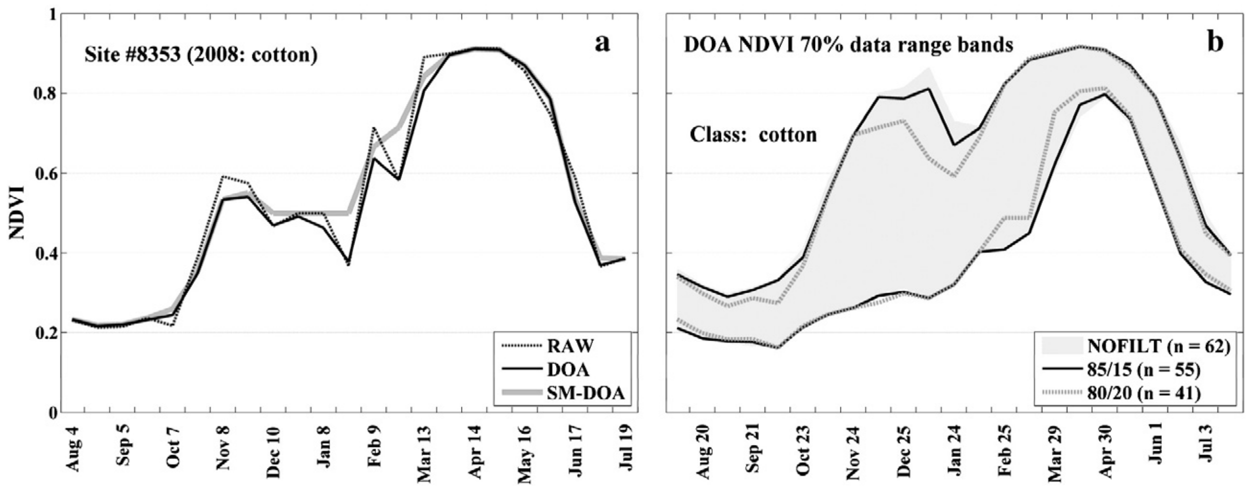

Fig. 3. (a) NDVI profiles for one cotton field site sample, showing variability from the three different levels of signal processing. This sample appeared in all three data sets with different levels of filtering applied. (b) DOA NDVI 70% data bands for cotton, illustrating the effects of the three different levels of filtering on the spectral profile distribution from this class.

the effects of filtering on spectral profile distribution, Fig. 3(b) shows DOA NDVI data bands from all three filtering levels for the cotton class.

By this construction, more profiles were identified as outliers and excluded from the analysis using the 80/20 percentiles than the 85/15 percentiles. In all cases, 2005 saw the most samples drop out, amounting to a loss of 22% of its points in the 85/15 dataset and 40%–41% of its points in the 80/20 dataset (Table 4). In the 85/15 dataset, years 2006–2009 lost similar proportions of data (9%–13%). In the 80/20 dataset, 2006 lost a greater proportion of its points (31%) than 2007–2009, with these years experiencing losses ranging from 23% to 26%. The fact that earlier years tended to lose more points suggests that more of these records might have been in error, reflecting increasing inaccuracy of farmer recollections going further back in time. A similar recency effect was observed in Rowley et al. (2007), which involved a survey that asked cattle ranchers to recall their grazing land productivity from past years.

4. Results and discussion

Averaged across the nine different data treatments (three levels of VI processing crossed with three levels of data filtering) and applied to the three different classification schemes (5-class, 4-class, and 2-class), NDVI classification percentage accuracy was 0.1% better than EVI using the CVYR evaluation method. The 5-year standard deviation of percent prediction accuracy averaged across the 27 scenarios (nine treatments and three classification schemes) was 0.3% lower using NDVI than EVI. Due to these small overall differences between NDVI and EVI performance, we present only results using NDVI in the interest of using the simpler VI.

Averaged across the three filtering levels and three classification schemes, the DOA NDVI dataset outperformed the RAW NDVI dataset by 1.9% and the SM-DOA NDVI dataset by 0.3%. The 5-year standard deviation of percent prediction accuracy averaged across the nine scenarios (three filtering levels and three classification schemes) was 0.7% lower using DOA NDVI compared to RAW NDVI and 0.3% lower than using SM-DOA NDVI. Thus we further limit the discussion to include only results obtained using the DOA NDVI dataset.

4.1. Effects of data filtering

As previously described, three different levels of point filtering (including no filtering) were applied to the DOA NDVI dataset. In this section, we compare accuracy results from the unfiltered (NOFILT) dataset to the 85/15 and 80/20 filtered datasets using the CVYR evaluation method.

Tables 5–7 show the 5-class, 4-class, and 2-class confusion matrices, respectively, from the three filtering levels, along with other accuracy statistics. In all three classification structures, overall accuracy improved as filtering became more stringent (5-class: 68.6% (NOFILT) vs. 79.3% (80/20); 4-class: 71.4% (NOFILT) vs. 80.9% (80/20); 2-class: 78.8% (NOFILT) vs. 84.3% (80/20)). In the 5-class design, the Kappa value (estimated using the $\hat{K}$ statistic; Congalton & Green, 1999) improved from 0.49 (NOFILT) to 0.66 (80/20). In the 4-class design, Kappa increased from 0.50 to 0.67, and in the 2-class design, Kappa increased from 0.57 to 0.68. Estimating the variance of $\hat{K}$ and converting $\hat{K}$ to a Z-score (Congalton & Green, 1999), classifications from the three filtering levels were all mutually distinct ($\alpha<0.01$) when 5 classes were

Table 4
Sample dropout results applying filtering to NDVI DOA data.

Filter	Data year	5-class	Dropouts	Percent	4-class	Dropouts	Percent	2-class	Dropouts	Percent
85/15	2005	284	80	22%	283	81	22%	261	74	22%
	2006	342	37	10%	330	49	13%	303	40	12%
	2007	324	33	9%	318	39	11%	290	34	10%
	2008	341	43	11%	339	45	12%	305	39	11%
	2009	332	47	12%	328	51	13%	295	46	13%
80/20	2005	218	146	40%	214	150	41%	199	136	41%
	2006	260	119	31%	260	119	31%	238	105	31%
	2007	272	85	24%	268	89	25%	248	76	23%
	2008	295	89	23%	297	87	23%	265	79	23%
	2009	282	97	26%	288	91	24%	258	83	24%

Table 5
Confusion matrices and accuracy statistics for the 5-class design.

		Reference class							
		Soy-Com	Soy-Cov	Soy-Cot	Cotton	Pasture	Total	User's	Producer's
NOFILT	Soy-Com	**599**	180	48	19	7	853	70%	74%
	Soy-Cov	180	**569**	12	17	41	819	69%	72%
	Soy-Cot	16	1	**22**	3		42	52%	24%
	Cotton	4	9	9	**22**		44	50%	35%
	Pasture	10	26	2	1	**66**	105	63%	58%
	Total	809	785	93	62	114	**1863**	**Overall accuracy: 68.6%**	
								Kappa = 0.49	
85/15	Soy-Com	**534**	130	36	12	2	714	75%	77%
	Soy-Cov	147	**547**	9	7	39	749	73%	79%
	Soy-Cot	13	2	**30**	3		48	63%	34%
	Cotton	2	3	11	**32**	2	50	64%	58%
	Pasture		14	1	1	**46**	62	74%	52%
	Total	696	696	87	55	89	**1623**	**Overall accuracy: 73.3%**	
								Kappa = 0.56	
80/20	Soy-Com	**483**	85	30	9	2	609	79%	84%
	Soy-Cov	84	**468**	2	2	30	586	80%	83%
	Soy-Cot	6	1	**29**	2		38	76%	43%
	Cotton		2	5	**27**	1	35	77%	66%
	Pasture	3	9	1	1	**45**	59	76%	58%
	Total	576	565	67	41	78	**1327**	**Overall accuracy: 79.3%**	
								Kappa = 0.66	

used. The same was true for the 4-class design, except that 85/15 and 80/20 were distinct only to the level of $\alpha<0.02$. In the 2-class design, only NOFILT and 80/20 were distinct with $\alpha<0.01$. 85/15 and 80/20 were distinct to the level of $\alpha<0.02$, while NOFILT and 85/15 were found to be indistinct ($\alpha>0.40$).

In each classification structure and for each class, User's Accuracy improved as filtering became more stringent. User's Accuracy indicates the likelihood that a user visiting a site on the ground assigned to a particular class by the model will find it in that state. The most substantial gains in the 5-class and 4-class designs occurred in the cotton class, where User's Accuracy jumped from 50% to 77% and 55% to 89%, respectively, as filtering stringency increased. This gain is attributable to the fact that a substantial number of cotton samples had a large, early-season (approximately Nov–Jan) NDVI "hump" that does not reflect cotton phenology but increases confusion (distributional overlap) with the other classes, which generally exhibit actual vegetative growth activity (and thus elevated NDVI) during all or part of this time window (see Fig. 2). The filtering method screened out many samples exhibiting this anomalous behavior. Considering that cotton is a later-season crop in Mato Grosso, a likely explanation for this occurrence of inflated, pre-cotton crop NDVI values is abundant and prolonged weed growth on the field sites in question prior to field preparation for cotton planting. Similar problems were found in Wardlow et al. (2006), where pre-crop weed growth produced NDVI increases that confounded comparisons between crop emergence timing and green-up onset in the state of Kansas in the U.S.

Producer's Accuracy also generally improved with filtering. Producer's Accuracy indicates the likelihood that a sample from a particular ground reference class will be classified correctly by the model. Values were typically higher than User's Accuracy for the larger classes and lower than User's Accuracy for the smaller classes. This behavior can be expected when (*i*) training class sizes are severely imbalanced, and (*ii*) overall accuracy is the value to be maximized during model optimization. Both features characterize the 5-class and 4-class

Table 6
Confusion matrices and accuracy statistics for the 4-class design.

		Reference class						
		Soy-Com	Soy-Cov	Cotton	Pasture	Total	User's	Producer's
NOFILT	Soy-Com	**713**	210	34	12	969	74%	79%
	Soy-Cov	177	**545**	17	40	779	70%	69%
	Cotton	7	3	**12**		22	55%	19%
	Pasture	5	27		**61**	93	66%	54%
	Total	902	785	63	113	**1863**	**Overall accuracy: 71.4%**	
							Kappa = 0.50	
85/15	Soy-Com	**626**	143	19	3	791	79%	83%
	Soy-Cov	127	**536**	6	38	707	76%	77%
	Cotton	3	3	**28**		34	82%	51%
	Pasture	2	14	2	**48**	66	73%	54%
	Total	758	696	55	89	**1598**	**Overall accuracy: 77.5%**	
							Kappa = 0.60	
80/20	Soy-Com	**541**	101	8	3	653	83%	84%
	Soy-Cov	97	**448**	1	23	569	79%	79%
	Cotton	3	1	**32**		36	89%	78%
	Pasture	2	15		**52**	69	75%	67%
	Total	643	565	41	78	**1327**	**Overall accuracy: 80.9%**	
							Kappa = 0.67	

Table 7
Confusion matrices and accuracy statistics for the 2-class design.

		Reference class				
		Soy-Com	Soy-Cov	Total	User's	Producer's
NOFILT	Soy-Com	**746**	201	947	79%	83%
	Soy-Cov	156	**584**	740	79%	74%
	Total	902	785	**1687**	**Overall accuracy: 78.8%** **Kappa = 0.57**	
85/15	Soy-Com	**628**	158	786	80%	83%
	Soy-Cov	130	**538**	668	81%	77%
	Total	758	696	**1454**	**Overall accuracy: 80.2%** **Kappa = 0.60**	
80/20	Soy-Com	**560**	107	667	84%	87%
	Soy-Cov	83	**458**	541	85%	81%
	Total	643	565	**1208**	**Overall accuracy: 84.3%** **Kappa = 0.68**	

studies. In this situation, a model often will have more to gain by expanding the reach of the larger classes at the expense of the smaller classes. Generally speaking, as the model strives to reduce errors of omission in the larger classes, this leads to a simultaneous (but necessarily smaller) increase in errors of commission in the larger classes. This behavior can result in an increase in errors of omission among the smaller classes while not necessarily reducing errors of commission within these classes to the same degree. To illustrate this point, notice that the 5-class, NOFILT User's Accuracy values range from 50% to 70%, whereas the Producer's Accuracy values range from 24% to 74%. This two-sided increase in Producer's Accuracy range compared to User's Accuracy range is seen across all three filtering levels and all three classification schemes, even in the 2-class results where there is less imbalance between the sample sizes (Tables 5–7).

In the 5-class design, the Soy–Cotton class had the lowest Producer's Accuracy value regardless of filtering level, ranging from 24% (NOFILT) to 43% (80/20). This class also had the lowest or nearly the lowest User's Accuracy values, ranging from 52% (NOFILT) to 76% (80/20). Absorbing this class into Soy-Com (the 4-class design) caused a slight deterioration (3%–4%) in Producer's Accuracy for the Soy-Com class at all three filtering levels, while User's Accuracy values increased (73% to 76% for 85/15) or were nearly unchanged (±1% for NOFILT and 80/20).

Among the other pairwise class comparisons, substantial confusion between Soy-Cov and Pasture was seen across all 5- and 4-class scenarios. This outcome is similar to confusion observed between single crop soybeans and pasture noted in Morton et al. (2006), which prompted those authors to implement a crop trajectory correction scheme to mitigate the problem. In our analysis, filtering helped to generally improve User's Accuracy for the Pasture class, but Producer's Accuracy did not improve from the NOFILT case except in the 4-class 80/20 evaluation.

Because the two largest classes (Soy-Com and Soy-Cov) comprise at least 90% of the testing samples in all of the considered filtering scenarios, it is important to examine the effects of filtering on these classes. In the 5-class design, User's and Producer's Accuracy values improve 9%–11% comparing NOFILT to 80/20, attaining at least 79% for Soy-Com and Soy-Cov under 80/20 filtering. Corresponding gains are similar in the 4-class design, with the exception that Producer's Accuracy for Soy-Com increases just 5%. However, the final 80/20 value in this regard is 84%, which is a generally favorable result none the less. In the 2-class design, User's and Producer's Accuracies improve 5%–7% when comparing NOFILT to 80/20, with all four 80/20 values in the range of 81% to 87%.

4.2. Estimating model accuracy

Three distinct accuracy assessments (CVYR, CV20, and INSAMP) were applied to the DT models for the different filtering levels and classification scenarios. Results are shown in Table 8. Difference values between error estimates are shown in the bottom three rows of the table. Despite efforts to reduce model overfit of the training data by using a 1% CF, there is still a substantial effect in this regard that is evident in all cases considered. In the 4- and 5-classs scenarios, INSAMP exceeds CVYR by 12.9–18.9% under the different filtering conditions. INSAMP achieves a maximum value of 98.1% overall accuracy in the 4-class design using 80/20 filtering, compared to 80.9% using the CVYR evaluation method. The smallest difference (6.4%) between INSAMP and CVYR is observed in the 2-class NOFILT scenario, though this value increases to more than 12% when filtering is applied. The average difference between INSAMP and CVYR across all nine cases is 14.2%.

The expected bias of CV20 compared to CVYR is evident as well. The maximum observed differences are 4.3% (4-class NOFILT) and 3.9% (5-class NOFILT), whereas three of the nine scenarios were found to differ by less than 1% (2-class 80/20 at 0.8%, 2-class NOFILT at 0.6%, and 5-class 80/20 at 0.5%). The average difference between CV20 and CVYR across all nine cases is 2.2%. Comparing these difference values to ‘std(CV20)’, in four of the nine cases we find that CVYR values are within a standard deviation band of observed CV20 values. Looking at the minimum observed CV20 value across the 30 trials, this value falls below CVYR in seven of the nine cases. Alternatively, if we compute the standard error about the mean (‘stder(CV20)’), we find that CVYR falls below this confidence band in all nine cases and in six of nine cases if we use a band twice this wide, indicative of statistical distinction between CV20 and CVYR and therefore bias of CV20 with respect to out-year classification.

Table 8
Model error estimates. ‘std’ denotes the standard deviation, and ‘stder’ denotes the standard error about the mean.

	5-class			4-class			2-class		
Statistic	NOFILT	85/15	80/20	NOFILT	85/15	80/20	NOFILT	85/15	80/20
min(CV20)	69.7	72.9	76.8	72.7	75.9	78.9	75.4	79.7	81.0
max(CV20)	77.5	80.0	83.0	79.4	84.4	86.4	82.5	86.9	90.1
avg(CV20)	**72.5**	**76.8**	**79.8**	**75.7**	**79.4**	**82.7**	**79.5**	**82.6**	**85.1**
std(CV20)	1.7	2.0	1.7	1.9	2.2	1.8	1.7	1.6	2.0
stder(CV20)	0.3	0.4	0.3	0.4	0.4	0.3	0.3	0.3	0.4
CVYR	**68.6**	**73.3**	**79.3**	**71.4**	**77.5**	**80.9**	**78.8**	**80.2**	**84.3**
INSAMP	**86.5**	**86.2**	**95.3**	**90.3**	**91.7**	**98.1**	**85.2**	**92.3**	**96.9**
INSAMP-CVYR	17.9	12.9	16.0	18.9	14.2	17.2	6.4	12.1	12.6
INSAMP-CV20	14.0	9.4	15.5	14.6	12.3	15.4	5.7	9.8	11.8
CV20-CVYR	3.9	3.5	0.5	4.3	2.0	1.9	0.6	2.4	0.8

4.3. Yearly results

Using a yearly hold-out method like CVYR allows us to examine variability in year-to-year performance of the DT models. Annual results for the different classification designs and filtering levels are shown in Table 9. The first thing to notice is the steady improvement of model performance across time, which is evident from the bottom row of values in the table. Averaged across all designs and filtering levels, year-on-year improvement ranges from 1.5%–2.5%. A convenient explanation for this occurrence is the recency effect observed earlier in the annual filtered sample dropout numbers, which suggests that the accuracy of the ground reference data degrades as one goes backward in time, due to limitations of farmer records and recollections. Likely some of these differences are also attributable to annual variations in growing season characteristics caused by weather. [As noted earlier, 2005 saw the greatest drop out of data with the filtering approach employed. It was the earliest year of the study, but it was also a year considered to suffer the most severe drought in the last 40 years (Hopkin, 2005), during which EVI green-up appeared noticeably affected (Samanta et al., 2010). This drought's effect on crops in 2005 is likely reflected in the MODIS data, as demonstrated by Asner and Alencar (2010)]. Another observation that can be gleaned from the 'StdDev' column in Table 9 is that data filtering reduces year-to-year variability in model performance, in addition to increasing overall accuracy.

4.4. Tree size and band usage

Through the error comparisons described earlier, we presented evidence that the constructed DT models overfit the training data to some degree despite efforts to mitigate this problem through our choice of a minimal CF value. Overfitting is often the consequence of a model having too many parameters relative to the number of data points (i.e., the model is overly complex). In DT modeling, each split is equivalent to a model parameter, so the total number of splits can be used as a measure of DT complexity.

Table 10 shows the year-by-year average tree sizes observed during the CVYR exercise. For the 5-class and 4-class designs, each one-year hold-out model was constructed using approximately 1500 training samples for NOFILT, 1300 samples for 85/15, and 1050 samples for 80/20. For the 2-class problem, these numbers drop to 1350, 1100, and 950, respectively. Average tree sizes, on the other hand, range from 7 splits (2-class NOFILT in 2005) to 98 splits (5-class NOFILT in 2005). The 5-class NOFILT scenario clearly was exceptional in that the five largest average tree size values occurred here. Of the remaining 5-class and 4-class cases, annual average tree size ranged from 19 to 36, and the median average tree size across all five years was fairly stable (28–32). Tree size dropped substantially with the 2-class cases, where annual average tree size ranged from 7 to 23, and the median average tree size across all five years was 12 (NOFILT, 80/20) and 16 (85/15). These smaller tree sizes suggest that the 2-class models were more parsimonious, which helps explain why the smallest bias values at each of the three filtering levels were observed with the 2-class models (Table 8; INSAMP-CVYR). However, this outcome is also partly attributable to the generally higher CVYR accuracies observed with the 2-class design, which limits the possible bias magnitude.

Table 11 shows band usage frequency, by classification design and filtering level. If a band (MODIS time period) appeared at least once in any of the 10 component trees of the boosted DT model, then it was deemed to be used. From Table 11, it is clear that band usage was fairly consistent and logical across all modeling scenarios. Specifically, we see information from the soy cropping time periods (roughly 6–14) and information from the peak of the second commercial, cover, or weed crop (roughly 17–19) used with the most frequency. On the other hand, information from the off-season (roughly 1–5, 20–23) was used with the least frequency.

Table 10
Hold-out year tree size (number of splits, averaged across 10 boosted DTs).

	Data	2005	2006	2007	2008	2009	Median
5-class	NOFILT	98	89	40	48	51	51
	85/15	26	29	30	32	30	30
	80/20	29	24	31	36	21	29
4-class	NOFILT	32	34	28	23	19	28
	85/15	29	32	20	34	33	32
	80/20	33	30	23	29	23	29
2-class	NOFILT	7	12	12	14	14	12
	85/15	15	22	22	15	16	16
	80/20	23	12	17	12	12	12
	Median	29	29	23	29	21	–

5. Conclusion

This research was made possible due to the 5-year ground reference data set collected by the authors (Coutinho and Victoria), which is unprecedented in spatiotemporal magnitude. With approximately 120,000 ha of agricultural land situated in the most densely cropped areas of Mato Grosso represented each year, these data provide an excellent means for developing and evaluating land-cover classification models as well as estimating potential accuracies that might be achieved under various agricultural classification designs. However, the field sites are concentrated in particular portions of Mato Grosso, and thus one can expect some drop-off in accuracy if applying a model constructed using these data to a larger area (such as the entire state). Likewise, one can also expect some accuracy decline if applying such a model to data from years not represented in the dataset.

The field dataset originally comprised 19 vegetative land-cover classes, most of which related to agriculture. A simple sample-size threshold was applied to remove poorly represented classes, and the smallest remaining class following threshold application (rice) was also removed

Table 9
Year-by-year CVYR overall accuracy values.

	Data	2005	2006	2007	2008	2009	Average	StdDev
5-class	NOFILT	60.2	66.0	71.4	72.4	72.8	68.6	5.4
	85/15	69.0	71.9	71.3	76.0	77.4	73.1	3.5
	80/20	75.7	76.5	77.2	81.7	84.0	79.0	3.6
4-class	NOFILT	66.2	69.1	73.1	71.6	77.0	71.4	4.1
	85/15	73.9	75.5	76.4	78.8	82.3	77.4	3.3
	80/20	79.4	78.5	76.5	84.9	84.0	80.7	3.6
2-class	NOFILT	71.0	78.4	79.6	82.0	83.0	78.8	4.7
	85/15	78.5	78.2	80.3	80.3	83.4	80.2	2.1
	80/20	80.4	82.4	84.3	86.4	86.8	84.1	2.7
	Average	72.7	75.2	76.7	79.3	81.2	–	–

Table 11
Band usage by classification design and filter level.

	Data	High usage (100%)	Low usage (<75%)
5-class	NOFILT	9–14, 17–19	1–5, 20–22
	85/15	11–14, 17–19	1–5, 7, 20–22
	80/20	10–14, 17–19	1–5, 7, 20–23
4-class	NOFILT	6–14, 17–18	2–5, 20–22
	85/15	6–14, 17–19	2–5, 20–22
	80/20	8–14, 17–18	2–4, 7, 20–22
2-class	NOFILT	8–14, 17–18	2–5, 20–22
	85/15	6–14, 17–19	2–5, 20–23
	80/20	9–14, 17–19	2–5, 7, 20–22

due to high VI statistical variability attributable to widely variable management of that crop. The 11 classes that remained generally can be categorized as pasture, single crop, single crop followed by cover crop, and double crop. The crop types and rotations represented in these 11 classes are fairly comprehensive for the region, accounting for 91.5% of the agricultural planted area found in Mato Grosso during the study period (IBGE, 2011).

Ideally, one would like to be able to discern as many of these specific classes as possible, so that reliable, maximally detailed land cover maps can be developed for use in land change, agro-environmental monitoring, and climate studies. Spectral similarity between VI time series from some of the classes precluded full separation, so that it became necessary to group particular classes to form super-classes. Specifically, we developed and evaluated two major groupings, which include soy followed by a commercial crop (Soy-Com) and soy followed by a cover crop or fallow (Soy-Cov). These groupings have economic and agronomic relevance, and thus represent an evolution of past single- and double-crop groupings such as those used in Galford et al. (2008) and Brown et al. (2007a). These super-classes potentially could be used in a hierarchical classification scheme as researchers pursue the future challenge of more specific crop-type mapping.

Several different variations of the MODIS NDVI and EVI datasets were subjected to parallel analyses in an attempt to identify an optimal dataset to use for the classification of the field data. On average across multiple class groupings (5-, 4-, 2-class), VI data treatments (RAW, DOA, SM-DOA), and field data point screening stringencies (NOFILT, 85/15, 80/20), NDVI with date-of-acquisition interpolation and no smoothing was found to be the best-performing data. This result is desirable because NDVI is a simpler index than EVI, and it also illustrates the value of utilizing the MODIS date-of-acquisition information to make the VI time series more temporally precise.

At best, some early land cover change studies for Mato Grosso evaluated classification accuracy using testing data from a different year than the model training data (e.g. Galford et al., 2008; Morton et al., 2006). However, none of these studies performed the type of rigorous, multi-year evaluation presented here, which makes it difficult to apply the results of those earlier studies for multiyear analyses. We tested multiyear applicability under nine distinct classification designs, and across 5 years of evaluations; we found maximum differences in annual accuracy that ranged from 6.4% (2-class, 80/20) to 12.6% (5-class, NOFILT), thus indicating the importance of multiyear studies for robustness of analysis.

By design, our results tend toward best-case expectations, giving some idea of potential accuracies that might be achieved in a regional-scale agricultural mapping program centered on Mato Grosso and neighboring areas. Only with 80/20 sample filtering do we realize 5-year out-of-sample accuracies consistently near or above 80% and Kappa values above 0.60. It should be noted that sample filtering was performed using all 5 years of data at once (i.e., not in a CVYR framework), likely imparting a small favorable bias to the 85/15 and 80/20 results. Our choice to use boosted decision tree models is supported in the literature; the observed tree sizes were reasonable and the most-used variables were logical. Alternative methods such as random forests (Clark et al., 2010), however, might produce better results. Further examination toward this end is warranted, though our expectations for realizing substantial improvement are tempered.

Our field data spatially and temporally have unprecedented coverage compared to previous agricultural mapping efforts in Mato Grosso. Nonetheless, the dataset has limitations that should be recognized, especially if used in an operational agricultural monitoring program over a larger area. First, the data are concentrated in central and west-central Mato Grosso. The VI data from these sites do not exhibit great variation in growing season timing, though we know this situation changes dramatically as one moves away from the area. For example, soybean planting 1000 km to the north near Santarém, Pará, occurs approximately 80 days later than planting in our study area (Brown et al., 2007b; Rudorff et al., 2011). Wider area studies are needed, and they must take crop calendar variations into consideration. Second, field data collection was facilitated immensely by established contacts with APROSOJA (Soy Producers Association) in Mato Grosso, who were extremely helpful in locating research participants. This also meant, however, that most farmers interviewed were soy producers. The dataset thus possibly over-represents soy cropping systems and under-represents other important non-soy cropping strategies such as single- or double-crop cotton, where the accompanying crop is something other than soy. Third, data collection did not take into account what varieties of soy, corn, or cotton were grown (e.g. fast maturing vs. slow maturing). Recent field work in Mato Grosso by author Coutinho has left the impression that there is substantial variability in planting and maturing times in these crops within the same general area. This variability is the byproduct of factors such as seed availability, market conditions, soil (seed bed) characteristics, and weather, all of which influence individual farm management decisions. Variations in these decisions likely contribute to the variability observed in the MODIS data and the difficulty in reaching higher classification accuracies. Fourth, it bears repeating that the accuracies obtained are best-case scenarios based on “pure” pixels that are hand-selected from field interiors. Accuracies, thus, are likely to be lower when trying to classify datasets that include mixed pixels.

The MAPAGRI Project, underway as of this writing, will address the field data limitations noted above, and it builds on the work presented here. MAPAGRI is a 3-year effort that began in 2011, led by the Brazilian Agricultural Research Enterprise (EMBRAPA) in collaboration with Brazil's National Space Institute (INPE) and numerous university-based satellite remote sensing labs across the country, to standardize methodology, data-sharing, and visualization protocols for the launching of an eventual national agricultural mapping system (Esquerdo, 2011). With additional and ongoing farmer involvement through programs like MAPAGRI, the recency effect that we observed during filter evaluation can be eliminated with the development and maintenance of a current, accurate, long-term record of field site data. Such a historical archive will be invaluable for reference as satellite time series and associated agricultural land use analyses are extended to decadal scales in the future. Crop types and management practices will no doubt continue to change, with changes in the global agricultural system, further highlighting the need for ongoing monitoring and ground data collection to keep up with Mato Grosso's rapidly changing agricultural landscape.

At the time of this writing, the authors were in the process of developing a time series of statewide land cover maps for Mato Grosso covering the 2001–2012 crop years that utilizes the ground reference dataset featured in this study along with other datasets. These maps will be used to examine land cover change dynamics across the state with a focus on agricultural intensification. Upon publication of those results, both the ground reference dataset and the map series will be made available to the public.

Acknowledgments

Authors Brown, Kastens and Bishop recognize the support of the College of Liberal Arts and Sciences, University of Kansas, the National Science Foundation under Award No. EPS-0903806, and matching support from the State of Kansas through the Kansas Board of Regents. These authors also wish to recognize the Kansas Applied Remote Sensing Program (KARS) of the Kansas Biological Survey for being supportive of this work through the provision of staff expertise, space, and computing resources. Darin Grauberger and assistants at the Cartographic and GIS Services lab of the Department of Geography at the University of Kansas helped produce the maps. Authors Coutinho and Victoria were supported by the Brazilian Agricultural Research Corporation (Embrapa). Many thanks go to all the farmers in Mato Grosso and representatives of the APROSOJA cooperative who participated in and facilitated our fieldwork.

References

Anderson, L. O., Shimabukuro, Y. E., Defries, R. S., & Morton, D. (2005). Assessment of deforestation in near real time over the Brazilian Amazon using multitemporal fraction images derived from terra MODIS. *IEEE Geoscience and Remote Sensing Letters*, *2*(3), 315–318.

Arvor, D., Jonathan, M., Meirelles, M. S. P., & Dubreuil, V. (2008). Detecting outliers and asserting consistency in agriculture ground truth information by using temporal VI data from MODIS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XXXVII* (Part B7).

Asner, G. P., & Alencar, A. (2010). Drought impacts on the Amazon forest: the remote sensing perspective. *New Phytologist*, *187*(3), 569–578.

Barona, E., Ramankutty, N., Hyman, G., & Coomes, O. T. (2010). The role of pasture and soybean in deforestation of the Brazilian Amazon. *Environmental Research Letters*, *5*(2) (9pp.).

Boryan, C., Yang, Z., Mueller, R., & Craig, M. (2011). Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International*, *26*(5), 341–358.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees.* Belmont, CA: Wadsworth.

Brown, J. C., Jepson, W. E., Kastens, J. H., Wardlow, B. D., Lomas, J. M., & Price, K. P. (2007). Multitemporal, moderate-spatial-resolution remote sensing of modern agricultural production and land modification in the Brazilian Amazon. *Giscience & Remote Sensing*, *44*(2), 117–148.

Brown, J. C., Kastens, J., Wardlow, B., Jepson, W., Coutinho, A., Venturieri, A., et al. (2007). *Using MODIS to detect cropping frequency variation in mechanized agriculture in Amazonia.* Proceedings of the Brazilian Symposium of Remote Sensing, Florianópolis, Brasil, 21–26 abril 2007, INPE.

Clark, M. L., Aide, T. M., Grau, H. R., & Riner, G. (2010). A scalable approach to mapping annual land cover at 250 m using MODIS time series data: a case study in the Dry Chaco ecoregion of South America. *Remote Sensing of Environment*, *114*(1), 2816–2832.

Congalton, R. G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data: principles and practices.* New York: CRC Press.

Coutinho, A. C., Victoria, D. de C., Paz, A. R. da, Brown, J. C., & Kastens, J. (2011). Dynamics of agriculture in the soy production pole of the state of Mato Grosso. *Proceedings of the Brazilian Symposium of Remote Sensing, Curitiba, Brasil, 30 abril – 5 maio, 2011, INPE* (pp. 6128–6135).

DeFries, R. S., & Chan, J. C. W. (2000). Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing of Environment*, *74*(3), 503–515.

DeFries, R. S., Morton, D. C., van der Werf, G. R., Giglio, L., Collatz, G. J., Randerson, J. T., et al. (2008). Fire-related carbon emissions from land use transitions in southern Amazonia. *Geophysical Research Letters*, *35*(22), L22705.

Epiphanio, R. D. V., Formaggio, A. R., Rudorff, B. F. T., Maeda, E. E., & Luiz, A. J. B. (2010). Estimating soybean crop areas using spectral-temporal surfaces derived from MODIS images in Mato Grosso, Brazil. *Pesquisa Agropecuaria Brasileira*, *45*(1), 72–80.

Esquerdo, J. C. D. M. (2011). Embrapa Agricultural Informatics Geotechnology Laboratory, GEO Task Ag-0703 JECAM. ftp://ftp.earthobservations.org/documents/meetings/20110502_ag_cop/Mapagri_English.ppt (last accessed November 6, 2012).

Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, *61*(3), 399–409.

Friedl, M. A., Brodley, C. E., & Strahler, A. H. (1999). Maximizing land cover classification accuracies produced by decision trees at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(2), 969–977.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., et al. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, *83*(1–2), 287–302.

Galford, G. L., Melillo, J., Mustard, J. F., Cerri, C. E. P., & Cerri, C. C. (2010). The Amazon frontier of land-use change: croplands and consequences for greenhouse gas emissions. *Earth Interactions*, *14*(15) (24pp.).

Galford, G. L., Melillo, J. M., Kicklighter, D. W., Cronin, T. W., Cerri, C. E. P., Mustard, J. F., et al. (2010). Greenhouse gas emissions from alternative futures of deforestation and agricultural management in the southern Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(46), 19649–19654.

Galford, G. L., Melillo, J. M., Kicklighter, D. W., Mustard, J. F., Cronin, T. W., Cerri, C. E. P., et al. (2011). Historical carbon emissions and uptake from the agricultural frontier of the Brazilian Amazon. *Ecological Applications*, *21*(3), 750–763.

Galford, G. L., Mustard, J. F., Melillo, J., Gendrin, A., Cerri, C. C., & Cerri, C. E. P. (2008). Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. *Remote Sensing of Environment*, *112*(2), 576–587.

Hansen, M. C., Dubayah, R., & DeFries, R. (1996). Classification trees: an alternative to traditional land cover classifiers. *International Journal of Remote Sensing*, *17*(5), 1075–1081.

Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 national land-cover database for the United States. *Photogrammetric Engineering and Remote Sensing*, *70*(7), 829–840.

Hopkin, M. (2005). Amazon hit by worst drought for 40 years. Published online October 11, 2005. Nature. http://www.nature.com/news/2005/051010/full/news051010-8.html (Last accessed November 6, 2012).

IBGE (2011). Sistema IBGE de Recuperação Automática, Produção Agrícola Municipal. Instituto Brasileiro de Geografia e Estatística. (In Portuguese). http://www.sidra.ibge.gov.br/bda/ (Last accessed November 6, 2012).

Jasinski, E., Morton, D., DeFries, R., Shimabukuro, Y., Anderson, L., & Hansen, M. (2005). Physical landscape correlates of the expansion of mechanized agriculture in Mato Grosso, Brazil. *Earth Interactions*, *9*(1) (18pp.).

Jonathan, M., Arvor, D., Meirelles, M. S. P., & Dubreuil, V. (2008). Field-oriented assessment of agricultural crops through temporal segmentation of MODIS VI data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XXXVII* (Part B7).

Martinelli, L. A., Naylor, R., Vitousek, P. M., & Moutinho, P. (2010). Agriculture in Brazil: impacts, costs, and opportunities for a sustainable future. *Current Opinion in Environmental Sustainability*, *2*(5–6), 431–438.

McIver, D. K., & Friedl, M. A. (2001). Estimating pixel-scale land cover classification confidence using nonparametric machine learning methods. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(9), 1959–1968.

Morton, D. C., DeFries, R. S., Shimabukuro, Y. E., Anderson, L. O., Arai, E., Espirito-Santo, F. D., et al. (2006). Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(39), 14637–14641.

Morton, D. C., DeFries, R. S., Shimabukuro, Y. E., Anderson, L. O., Del Bon Espírito-Santo, F., Hansen, M., et al. (2005). Rapid assessment of annual deforestation in the Brazilian Amazon using MODIS data. *Earth Interactions*, *9*(8) (22pp.).

Richards, J. A., & Jia, X. (1999). *Remote sensing digital image analysis: an introduction* (3rd ed.). Heidelberg: Springer-Verlag.

Rowley, R. J., Price, K. P., & Kastens, J. H. (2007). Remote sensing and the rancher. *Rangeland Ecology & Management*, *60*(4), 359–368.

Rudorff, B. F. T., Adami, M., Aguiar, D. A., Moreira, M. A., Mello, M. P., Fabiani, L., et al. (2011). The soy moratorium in the Amazon biome monitored by remote sensing images. *Remote Sensing*, *3*(1), 185–202.

RuleQuest Research (2012). Description of boosting used in See5. http://www.rulequest.com/see5-win.html#BOOSTING (Last accessed November 6, 2012).

Samanta, A., Ganguly, S., Hashimoto, H., Devadiga, S., Vermote, E., Knyazikhin, Y., et al. (2010). Amazon forests did not green-up during the 2005 drought. *Geophysical Research Letters*, *37*(L05401) (5pp.).

Soybean & Corn Advisor, Inc. (2009–2011), Cordonnier, Michael. May 5, 2011. http://www.soybeansandcorn.com/news/May5_11-Sorghum-and-Millet-are-Alternatives-to-Safrinha-Corn-Production(Last accessed November 6, 2012).

Wardlow, B. D., & Egbert, S. L. (2007). Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the U.S. Central Great Plains. *Remote Sensing of Environment*, *112*(3), 1096–1116.

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment*, *108*(3), 290–310.

Wardlow, B. D., Kastens, J. H., & Egbert, S. L. (2006). Using USDA crop progress data for the evaluation of greenup onset date calculated from MODIS 250-meter data. *Photogrammetric Engineering and Remote Sensing*, *72*(11), 1225–1234.